Quality-Based Framework for Requirement Analysis in Data Warehouse

Munawar

Informatics Department
Esa Unggul University
Jakarta, Indonesia
an moenawar@yahoo.com

Naomie Salim, Roliana Ibrahim Information System Department Universiti Teknologi Malaysia Johor Bahru, Malaysia (naomie,roliana)@utm.my

Abstract—Data warehouse (DW) integrates data from external and internal sources not only for storage purposes but also for analythical query processing. As this information is to be used for business intelligent, it makes the design of DW difficult task. Requirements analysis is the foundation for the development of a DW. Ensuring quality in requirements analysis is an importance task to minimize the risk of failure in a DW project. However, quality issues in DW requirements analysis have not given much attention in the literature. In this paper we have advocated the integration of quality and multi-driven approach (goal-driven, user-driven, data-driven, process-driven and external-driven) as an explicit stage in the DW development to reduce failure risk of DW.

Keywords—requirements analysis, data quality, quality integration.

I. INTRODUCTION

The data warehouse (DW) is a special database with many years of historical information to satisfy decision making request [10]. This historical information's are built based on a large amount of data collected from heterogeneous sources into a multi-dimensional model in order to facilitate repository and trend analysis.

The first stage of DW development deals with the requirements analysis. Requirements analysis for DW represents collections of relevant information for decision making process [11]. The user needs or the actual availability of data are major sorces of requirements analysis.

Successful DW needs to be based upon good requirements analysis. A good quality DW can be triggered by a good quality of requirements analysis. Unfortunately, requirement analysis is often unclear and uncertain because of changing the requirements can be happened even in the short-term [6] to accommodate the fast evolution of the business conditions. Consequently, poor requirements analysis has been found to be a major failure of many software projects (including DW) [27].

Due to the strategic importance of DWs, quality should be incorporated from the requirements analysis stage [24]. The lack of quality incorporation in requirements analysis has a major impact on the DW success [25].

This paper is extended concept of IRADAH method [15, 16], in order to incorporate quality into the requirements phase in DW development. Multi-driven is our approach which integrates five approaches: goal-driven, user-driven, data-driven, process-driven, and external-driven. Besides satisfies user expectations, quality integration issues in the early stage of DW development is also incorporated to minimize the risk of failure in DW projects.

The paper is organized as follows. In the next section, related works is discussed. Section 3 proposes the framework to incorporate quality into the requirements analysis of DW design. Section 4 explains the interpretation of proposed framework in a real case study. And finally, we conclude in the last section.

II. RELATED WORKS

Ensuring quality is an important task to perform in a DW and should be done from the early stage of DW development. Quality of conceptual, logical, and physical have been done by many authors. However, to the best of our knowledge there is no significant work at requirements level can be seen in the DW literature.

Information scenario is presented by [22] to identify the DW requirements using a Goal-Decision-Information (GDI) diagram. However the authors did not explain the correlation between information scenarios and requirements.

Authors of [25] proposed an interesting approach denoted easyREMOTEDWH (easy Requirements Modeling Technique for Data Warehouses) to represent business objectives and needs from different stakeholders' perspectives. Unfortunately, there are no guidelines or any notation to properly specify the DW requirements..

A different approach is proposed by [3] that uses SOM (Semantic Object Model) to exploit the business process model for the DW requirements analysis. Focus of this approach is on the subject of the monitoring without explanation why monitoring is required.

According to [5], DW requirements can be obtained by exploting the requirements of business process and strategic decision process. DW models can be produced by combining DW requirements and as-is data model. New requiremets can

be elicited whenever the model is obtained. The study of [9] indicates that detailed business process analysis is not always feasible. Therefore it is not good to start requirements analysis with detailed analysis of business process

Our proposed frameworks mainly focus on integrating quality into the requirements analysis in designing a DW. Quality can be reached by puts together (data, goal, user, process, and external factor)-driven in this framework. Therefore any changing in requirements can be detected earlier.

III. PROPOSED FRAMEWORK

A. Requirements Analysis for Data Warehouse Development

Data warehouse is a complex system consisted of several sub systems and processes between them. A framework provides guidance to the designers of DW by linking the main components of DW architecture to a model of DQ that will be integrated. At the requirements analysis, the information needed to be maintained in the DW is discovered. The next phase, that is the conceptual design, can be said as abstraction of the users' request to some information structures, which act as the bridge connecting the real world and the machine world.

Requirements analysis is an important phase within DW project to minimize the risk of failure. Requirements analysis is the foundation for all project activities and influence most life cycle choices [13]. Nevertheless requirements analysis for DWs is often neglected and has not been paid attention [24; 31]. On average, requirements analysis failed in addressing the required information in DW development. Consequently, has a major impact on the success of DW project [25].

The requirements analysis for DWs is different than those which apply to other types of information systems. The approaches adapted for requirements analysis for DW development are (a) interview (b) workshop (c) prototyping (d) scenarios (e) subject area [7; 33]. There are several approaches how to determine the requirements for the development of DW.

- The data-driven (also called supply-driven) approach [34; 35] is a bottom-up technique that based on exploration of the data sources in order to identify all the available data. Specifically, data-driven approach tries to construct DW based only on actual data in operational system and disregarding business goals and user needs. Organization needs are not identified, or are identified only partly.
- 2) The process-driven approach identifies the most important business process that requires measurement and control [3;12] and then aligns them with corporate strategies. A DW is designed based on a set of relevant business subjects [3] and each of these subjects centers around a business process.
- 3) The goal-driven approach is typically top-down. It stresses on the alignment of DW with corporate strategy and business objectives [6] by interviewing the top-management. Different visions are then analyzed and merged in order to obtain a consistent picture and finally translated into quantifiable KPIs.

- 4) The user-driven approach is bottom-up technique that underline the involvement of end users in DW [36;13] to determine required information from different business users. Their opinions are then combined and made consistent in order to obtain a unique set of multidimensionl schemata.
- 5) The external-driven approach is the need for handling the pressure to comply with governmental regulations (such as bank central regulation for banking industry) and others external pressure (such as regulation for public listed company) which require a real-time disclosure about business operations [4]. This approach is typically top-down and the main reason behind the current growth rate of such DW initiatives.

More than one approach must be put to work to obtain a data model in avoidance the drawbacks of the single approaches, which reflects the analytical needs of an organization in a precise and appropriate way [2]. It also can be said that all of these approaches are complementary and should be used in parallel to achieve optimal design [29; 35]. We hereby describe a multi-driven and integrated methodology for developing DW model, which can be illustrated in Figure 1.

In summary, the strengths and weaknesses of requirements analysis techniques can be seen in Table I.

TABLE I. THE STRENGHTS AND WEAKNESSES OF REQUIREMENTS
ANALYSIS TECHNIQUES IN DATA WAREHUSE DEVELOPMENT

ANALYSIS TECHNIQUES IN DATA WAREHUSE DEVELOPMENT						
Requirements Analysis	Strenghts	Weaknesses				
User-driven	Involvement of end users is essentials in DW projects to ensure the successful use of DW [2]	Unclear of users' understanding of DW, business strategies or organizational processes, make the degree of obsolescence of resulting schemata is high [2] It takes time expensive to achieve concensus on requirements with many different point of view [2;1]				
Data-driven	 The fastest way to define a DW model [2] Simpler [32] Very stable [32] 	Such models perhaps do not reflect all of the facts that are needed in analysis business goals [2] User involvement is limited [32] Multidimensional schemata produced could not match with user requirements if information is not actually present in the data source [32]				

Req <mark>uire</mark> ments Analysis	Strenghts	Weakn <mark>es</mark> ses
Goal-driven	Correct identification of the relefant indicators can be obtained [2;1]	Very dependable to the participation of top management in requirements analysis process [1] Need high capable staff in translation process from the collected high level requirements into quantifiable KPIs [2;1] It is hard to predict the needs of all senior managements [2]
Process-driven	Essential business processes and indicators to measure these processes are identified [2;3] Close to the need of business and adaptive with the business environment [30]	The model reflect business processes not process of decision making [2]
External-driven	Comply with governmental regulations or others external pressure require a disclosure about business operations [4]	Sometimes external data is needed to generate required information

Main focus in requirements analysis for DW development is subject area. Subject area is the highest level of data model in a DW. Policies, procedures, and rules can be included into the subject area to gain the detail information taxonomy of a DW [1]. Detail explanation about Figure 1 can be read in [16]

B. Data Quality Dimensions in Requirements Analysis

In DW systems, requirements analysis forms the basis for DW design. The quality of DW depends on the quality of requirements gathered. It is essential to assure the quality of the DW as it became the main tool for strategic decisions. Besides understanding the business context, quality consideration is very much needed in the system being developed. The best time to begin integrating quality is before the DW is first developed [26], by determining the important data quality (DQ) dimensions [28]. Various DQ dimensions that should be taken into account in requirements analysis can be seen in Table II.

C. Quality-Based Framework for Requirements Analysis

The development of DWs is particularly complex. It needs many phases to develop a DW. DQ issues can arise at any stage of data warehousing. All these phases of data warehousing are responsible for DQ issues in the DW. Therefore the need a framework is essential in order to guarantee the DQ continuous improvement in the whole DW phase.

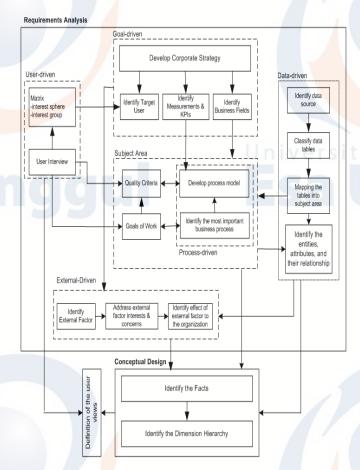


Figure 1. Requirements Analysis for Data Warehouse Development

. Effective business decision-making depends on good DQ, and poor DQ can be costly and sometimes disastrous [14]. A key factor in the success of DW is the quality of the data provided [13]. Organizations therefore need to understand DQ and establish procedures to asssure the DQ in DWs from the early process.

In a DW context, there are classification in requirements: functional and non-functional. Functional requirements deal with *what* data to be stored, while non-functional requirements specify *how* to provide information to facilitate reporting and analysis in a correct manner. Functional requirements can be: data-driven, goal-driven, user-driven, and external-driven, while process-driven can be non-functional requirements.

A scenario based design method is very useful to determine what functionalities are required for the system [8]. Scenarios can be used to systematically derive all the requirements. Varied situations can be easily exemplified using scenarios. Also, implementation of the DW becomes easier due to the clear structure of scenarios.

Decision in adoption between goal-driven and user-driven may become unclear. The goal-driven is top down, while the user-driven is bottom up. Progressive refinement is necessary in goal-driven based on the goals appointed by a the excutives. While in the user-driven, combining the closer requirements is needed to gain the requirements from the business-users'

perspectives. Unclear understanding about strategy and goals of organization will give a different result between goal-driven and user-driven approach.

TABLE II. DATA QUALITY DIMENSIONS IN REQUIREMENTS ANALYSIS

	D . O . W. Di I			Requirements Analysis					
	Data Quality Dimension		UD	GD	ED	DD	PD	ć	
	N.	Performance			1				
	Non Functional	Security			1	0			
		Multidimensi	onality		4				
Business		User Friendliness		4	1			1	
Quality		Represent Fa	et		1		1		
		Distinguish D	imensions to facts		1		1		
	Functional	Summarisabil	ity		4		1		
		Integration with Data Source		4	1		1	1	
		Fast Track o	f User Req. Changes		4		1		
		High Quality Documentation					1		
	Content	Soundness	Accuracy	1	4		1	1	
			Consistency	1	4	√	1	1	
			Applicability	1					
			Timeliness	1	1	1	1	1	
			Conciseness	1					
		Relevance	Comprehensiveness	4					
		Relevance	Clarity	4					
Information			Correctness	4			1		
Quality			Security	4					
		Infrastructure	Currency				1		
		IIII astroctore	Accessibility	1			1		
	Access		Convenience	1	V (a r	C	+	
			Traceability	1	4			1	
		Process	Maintainability		4				
			Interactivity	1	4				
			Speed	1	4			4	
	Correct						1		
Technical	Unambiguous						4		
Quality	Consistent	Consistent					1		
	Complete			1			1	1	

Legend UD = User-Driven DD = Data-Driven ED= External-Driven GD = Goal-Driven PD = Business Process-Driven

The Semantic Object Model (SOM) is a comprehensive and integrated methodology for business engineering [3]. It supports sound modeling of business systems and can be used for analysis and design.

Anticipation to the fast evolution within the related industry is the key strategy to win the competition. Sustainability of an organization depends on the ability of an organization to strengthen their competitive advantage [21]. Porter's five competitive forces [20] were selected because both academia and industry leaders worldwide accepted this tool.

The summary of framework for requirements analysis in DW design which incorporate quality can be seen in Table III.

D. Deliverables in Requirements Analysis

1) Fact

The DWcornerstone of development multidimensional modeling. A multidimensional model consist of facts and dimensions. Facts are central to DW. Analyzing user requirements implies identification of facts by perceiving the metrics behind user demands [7]. Facts are groupings of analysis indicators or measures in business process.

2) Preliminary Workload

The most important parts of the DW development is ETL processes [20] because they are responsible of integrating data from heterogeneous data sources into the DW repository. Besides time consuming, designing ETL processes is verymuch complex and costly [23]. UML behavior diagram (specifically activity diagram) can be used to show the activities involved in ETL processes together with the DW conceptual schema [18].

IV. PRACTICAL EVIDENCE

IRADAH method was applied to a students' admission project in a private university. During this process, scenario based design method was be used to derived all requirements. Varied situations can be easily exemplified using scenarios. Also, implementation of the DW becomes easier due to the clear structure of scenarios.

The external-driven determines information sources influencing students' admission and reward effect to attract the students candidate. The user-driven identifies the proper media for advertising and the actual availability of data to support this decision such as campaign events and its budget estimation. The goal-driven emphasis on the alignment between the DW and university strategy and its business objectives. In students' admission, the main goal is finding the best strategy to increase students intake with the proper budget. The data-driven exploits the students' admission data to generate information based on desired KPIs (Key Performance Indicators). But there is a risk of generating information by using computation if the data is not actually present or when the desired KPIs are not directly available. The process-driven identifies the students' admission process which is often cross-functional in the university. Gathering information related to admission process from different sources, monitoring this process, and aligning them with university strategies and high-level goals is a major issue for decision makers. Finally, the DFM can be adopted as a conceptual model enabled us to get dimensional fact model for students' admission as depicted in figure 2.

TABLE III. QUALITY-BASED FRAMEWORK FOR REQUIREMENTS ANALYSIS IN DATAWAREHOUSE DEVELOPMENT

DW Development Phase	Input	Processes	Quality Drivers	Tools	Deliverables
Requirements analysi	s → to discover the infor	mation needed to be maintained in the D	W		
✓ User-Driven	Job description	Stress involvement of end users in data warehousing Aids users on identifying their analytical needs	Non Functional Functional Information Access Information Content Technical	Interviews Workshops Prototyping Scenario Matrix interest sphere	Fact Focus of decision making process Preliminary workload
✓ Goal-Driven	Company profile Official Corporate document	Elicit the information to be kept in decision support To align DW with corporate strategy & business objectives	Information Content	Corporate strategy KPIs	A set of queries, expressed according to a high level language
✓ Data-Driven	Existing documentation Database schema	Reengineer data source into the logical data schema Reorganize the identified source schema to form DW models	Functional Technical	Entity Relationship (ER) Model or Class Diagram	Esa
✓ Process-Driven	Standard operating procedure Work flow processes	Conversion of data structures of business process into DW structure Finding data sources used by the process	Functional Information Content	SOM (Semantic Object Model)	
✓ External Factors- Driven	Government regulation related to organization Standard value for the same industry	Address external factor interest & concerns Identify effects of external factor to the organization	Information Content	Five Forces Porter [20]	
	to abstract the users' req	uest to some information structures, which			ine world
✓ Multidimensional Modeling	Fact Preliminary workload	Identifying the fact of interest Identifying the dimension hierarchies Identifying measure Identifying aggregation	Functional Information Content	ME/R ER Model DFM (Dimensional Fact Model)	Dimensional scheme Dimensional scheme is designed to store data in a way that: • Emphasized understandability • Enhances query performance • Accommodate change

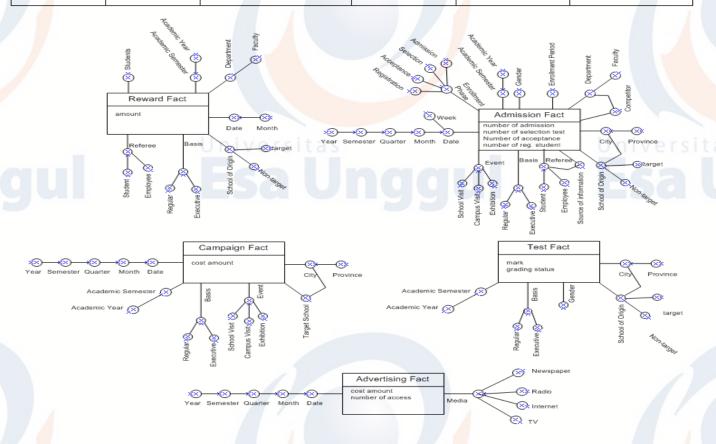


Figure 2. Dimensional Fact Model for Students Admission

V. CONCLUSIONS

The proposed extension in the IRADAH method has been used to provide an approach for integrating data quality and multi-driven approach (goal-driven, user-driven, data-driven, process-driven, and external-driven) into the requirements analysis in DW development. This study establishes the need to ensure that quality aspects should be incorporated properly at the requirements level in the DW development in order to minimize risk of DW project failure. Based on the case study in students' admission, it indicates the importance of this framework in designing a DW. Unfortunately, this study is only limited to one case study. We need to test this model in several companies, with different area of businessess.

REFERENCES

- [1] Guo, Y., Tang, S., Tong, Y., & Yang, D. 2006. Triple-driven data modeling methodology in data warehousing: a case study. In *Proceedings ACM International Workshop on Data Warehousing and OLAP*, 59-66.
- [2] Niedrite, L., Treimanis, M., Solodovnikova, D., and Grundmane, L. 2009. Development of Data Warehouse Conceptual Models: Method Engineering Approach. IGI Global
- [3] Lujan-Mora, S., Trujillo, J., & Song, I. 2002. Extending the UML for multidimensional modeling. In *Proceedings of UML*, LNCS 2460, (pp. 290-304). Springer.
- [4] Frolick, M., & Ariyachandra, T. (2006). Critical Success Factors in Business Performance Management-Striving for Success. *Information Systems Management*, 25, 113-120.
- [5] Sapia, C., Blaschka, M., Hofling, G., & Dinter, B. 1999. Extending the E/R model for the multidimensional paradigm. Lecture Notes in Computer Science, 1552, 105-116.
- [6] Frendi, M., Salinesi, C., Requirements engineering for data warehousing, proceedings of REFSQ Workshop, 2003.
- [7] Giorgini, P., Rizzi, S., Garzetti, M.: GRAnD: A goal-oriented approach to requirement analysis in data warehouses. Decision Support Systems 45(1), 4–21 (2008)
- [8] Paim, F. R. S., and Castro, J. B. 2003. DWARF: An Approach for Requirements Definition and Management of Datawarehouse Systems. In Proceedings of the 11th IEEE International Conference On Requirement Engineering (pp 75-78)
- [9] Matteo Golfarelli, Dario Maio, Stefano Rizzi. 1998. Conceptual Design of data warehouses from E/R schemes. Proceedings of the Hawaii International Conference on System Sciences.
- [10] Hua, L Teck. 2011. Sustainable Competitive Advantage for Market Leadership amongst the Private Higher Education Institutes in Malaysia. 2nd Internation Conference on Business and Economic Research (2nd ICBER 2011) Proceeding
- [11] Jeusfield, M., Quix, C., Jarke, M. 1998. Design and Analysis of Quality Information for Data Warehouses. In proceedings of 17th International Conference on Conceptual Modeling (ER'98), Singapore.
- [12] Rizzi, S. 2009. Conceptual Modeling Solutions for the Data Warehouse. IGI Global
- [13] Kaldeich, C., & Oliveira, J. 2004. Data warehouse methodology: A process driven approach. In *Proceedings of CAISE*, LNCS, 3084, 536-549
- [14] Kimball, R., Reeves, L., Ross, M., and Thornthwaite, W. 1998. The Data Warehouse Lifecycle Toolkit. John Wiley & Sons, second edition.
- [15] Wang, R.Y., and Strong, D.M., 1996, Beyond accuracy: What data quality means to data consumers, *Journal of Management Information* Systems. 12(4): 5-34.
- [16] Munawar, Naomie Salim, and Roliana Ibrahim. 2011. Toward Data Quality Integration into the Data Warehouse Development. Ninth IEEE International Conference on Dependable, Autonomic and Secure

- Computing. 978-0-7695-4612-4/11 © 2011 IEEE Computer Society. DOI 10.1109/DASC.2011.194
- [17] Munawar, Naomie Salim, and Roliana Ibrahim. 2011. Toward Data Warehouse Quality through Integrated Requirements Analysis. ICACSIS 2011. ISBN: 978-979-1421-11-9.
- [18] Munawar, Naomie Salim, and Roliana Ibrahim. 2012. Comparative Study of Quality Dimensions for Data Warehouse Development: A Survey. A. Ell Hassanien et al. (Eds.): AMLTA 2012, CCIS 322, pp. 465–473, 2012. © Springer-Verlag Berlin Heidelberg 2012
- [19] Munoz, L., Mazon, J., Pardillo, J., and Trujillo, J. Modelling ETL Processes of Data Warehouses with UML Activity Diagrams in R. Meersman, Z. Tari, and P. Herrero (Eds.): OTM 2008 Workshops, LNCS 5333, pp. 44–53, 2008. c_Springer-Verlag Berlin Heidelberg 2008
- [20] Vaisman, A.2006. Requirements Elicitation for Decision Support Systems: A Data Quality Approach. Proceedings of 8th International Conference on Enterprise Information Systems May 23-27, (2006).
- [21] Porter, M.E. 2008. The Five Competitive Forces that Shape Strategy. Harvard Business Review. January, 2008.
- [22] Porter, M.E. 1985. Competitive Advantage: Creating and Sustaining Superior Performance (1st ed.).New York: The Free Press.
- [23] Prakash N., Singh Y., Gosain A., "Information scenarios for data warehouse requirements elicitation". 23th International Conference on Conceptual Modelling (ER2004). Volume 3288 of LNCS, pages 205-216. Springer-Verlag (2004).
- [24] Rainardi, V. 2008. Building a Data Warehouse: With Examples in SQL Server. Apress. ISBN-13 (electronic): 978-1-4302-0527-2
- [25] Rizzi, S., Abelló, A., Lechtenbörger, J., & Trujillo, J. 2006. Research in data warehouse modeling and design: Dead or alive? In *Proceedings of the 9th ACM Int. Workshop on Data* Warehousing and OLAP (DOLAP '06), (pp. 3-10) ACM Press
- [26] Schiefer J., List B., Bruckner R. 2002 A holistic approach for managing requirements of data warehouse systems. In Proceeding of Americas Conference on Information Systems.
- [27] Thomann, J and Wells, D. 1999. Data warehouse quality management.
 The Data Warehousing Institute's Fourth Annual Implementation
 Conference, held February 14-19, 1999 in Anaheim, CA
- [28] Thayer, R. H. 2002. Software Requirements Engineering Concepts in Software Engineering, vol.1: The Development Process, Richard H. Thayer and Merlin Dorfman, 2nd ed., IEEE Computer Society Press, Los Alamitos, CA.
- [29] Singh, R and Singh, K. 2010. A Descriptive Classification of Causes Data Quality Problems in Data Warehousing. IJCSI International Journal of Computer Science Issues. Vol. 7, Issue 3, No 2 May 2010. ISSN: 1694-0784
- [30] Guo, Y., Tang, S., Tong, Y., & Yang, D. 2006. Triple-driven data modeling methodology in data warehousing: a case study. In Proceedings ACM International Workshop on Data Warehousing and OLAP, 59-66.
- [31] Chowdhary, P., Mihaila, J., and Lei, H. 2006. Model Driven Data Warehousing for Business Performance Management. IEEE International Conference on e-Business Engineering (ICEBE'06). 0-7695-2645-4/06
- [32] Winter, R., Strauch, B. 2003. A Method for demand driven information requirements analysis in data warehouse projects. Proceeding of the Hawai International conference on system sciences. January 6-9
- [33] Golfarelli, M. 2010. From User Requirements to Conceptual Design in Data warehouse Design. IGI Global. DOI: 10.4018/978-1-60566-756-0.ch001
- [34] Sen, A., and Sinha, A. P. 2007. Toward Developing Data Warehousing Standards: An Ontology-Based Review of Existing Methodology. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS— PART C: APPLICATIONS AND REVIEWS, VOL. 37, NO. 1, JANUARY 2007
- [35] Artz, J. 2005. Data driven vs. metric driven data warehouse design. In Encyclopedia of Data Warehousing and Mining, (pp. 223 – 227). Idea Group

Universitas Esa Undau

- [36] List B., Bruckner R., Machaczek K., and Schiefer, J. 2002. A comparison of data warehouse development methodologies: Case study of the process warehouse. In *Proc. DEXA*
- [37] Westerman, P. 2001. Data Warehousing using the Wal-Mart model. (p. 297). Morgan Kaufmann

ggul

Esa Unggul

Esa U











